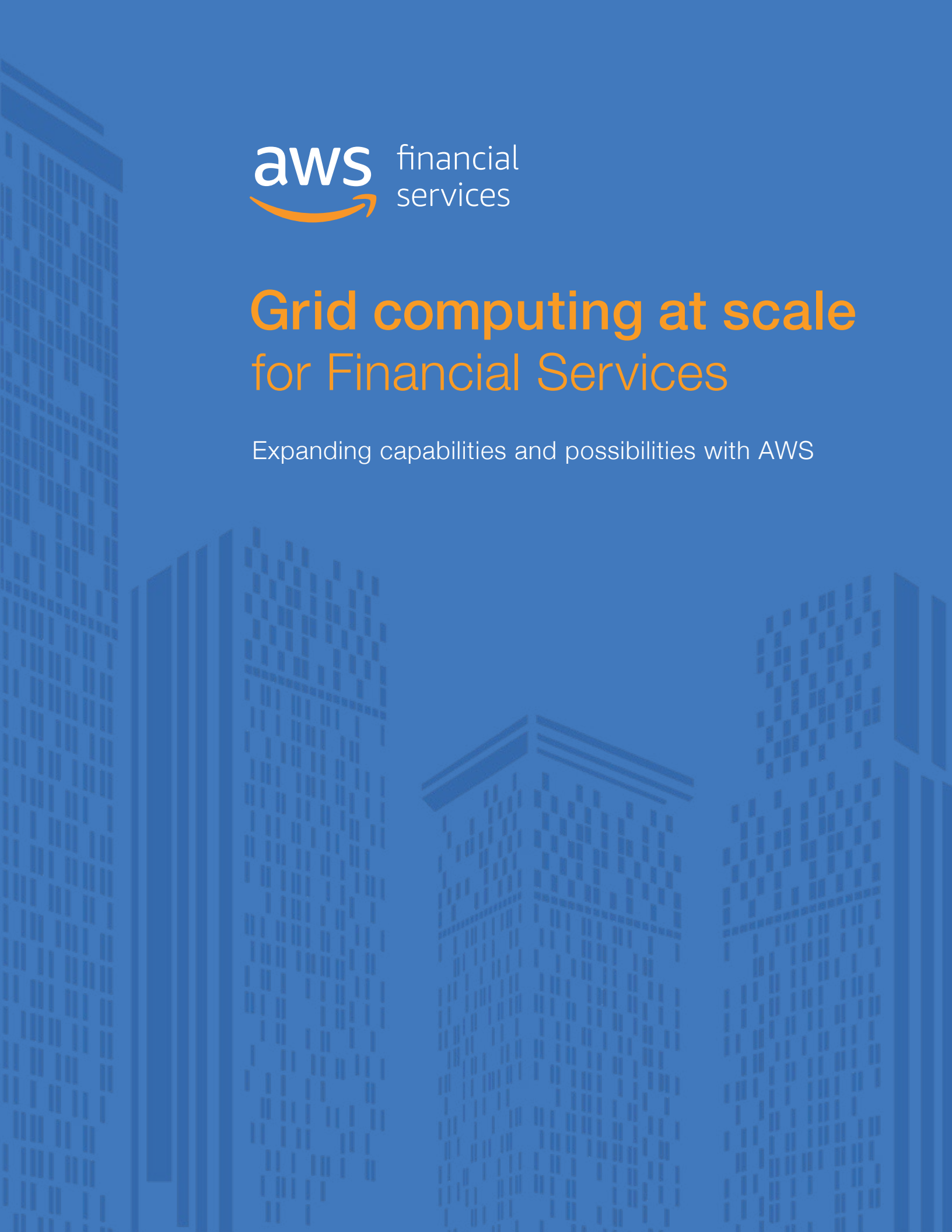


Grid computing at scale
for Financial Services

Expanding capabilities and possibilities with AWS

Grid computing in Financial Services today

What projects would you pursue if you were able to add 100,000 cores to your compute grid, shut them down when you were done, and only pay for them when you use them? How would you enhance your risk management, product development, quantitative research, or capital management processes?

The need to conduct compute-intensive calculations in areas such as risk management, regulatory compliance, product development and pricing, trading, clearing, and surveillance is not new to the Financial Services. However, the volume and timeliness of these calculations have increased dramatically. These changes, combined with heightened regulatory requirements and dynamic market conditions, are forcing companies to consider expanding their on-premise grid-computing capabilities—which can be capital-, labor-, and time-intensive. Instead, organizations are taking advantage of storage and computing capabilities that they can quickly and securely scale on demand. To address their grid-computing needs, financial institutions are increasingly turning to AWS and are realizing a range of benefits, including faster processing, lower total costs, and greater accessibility.

Familiar challenges at unfamiliar scale

Financial simulations are essential to the operations of Financial Services institutions in order to identify and manage risk, fully comprehend capital positions, and make informed investment and pricing decisions, among other uses. What has changed to make these simulations more onerous?

Regulatory bodies are requiring Financial Services institutions to perform increasingly stringent stress tests to maintain adequate capital ratios.

Obligatory stress-test requirements are further complicated by stipulations around using capital and collateral to maintain specified liquidity levels in order to conduct business.

The development of new quantitative trading strategies and complex products requires a greater variety of data sets and increased modelling capacity, which in turn adds to the complexity of design and back testing.

These new burdens, which are exacerbated by limited datacenter capacity, affect organizations across the Financial Services industry and can lead to significant delays in job completion. These delays can in turn deter users from running jobs altogether, which introduces a range of financial and operational risks.

FRTB: a case study in regulatory evolution and expanding computing needs

Working with regulators globally, the Basel Committee on Banking Supervision developed the Fundamental Review of the Trading Book (FRTB) regulation, which is now due to be implemented by January 1, 2022, following a revision. This regulation will require significant changes to the processes and infrastructure banks currently rely on to assess their market risk. FRTB requires the use of more granular risk factors and expands the range of scenarios and the amount of historic data that organizations will need to analyze to calculate their exposures. For instance, FRTB's requirement that banks use 10 years of historical data to run simulations will increase the number of calculations they need to make. Industry feedback indicates a three- to five-fold increase over current Basel requirements.

Limited in-house grid capacity

The challenge of building and maintaining an effective on-premise grid boils down to the difficulty of meeting a variable demand with a fixed resource. It is nearly impossible to correctly size a static grid that can effectively accommodate shifting daily computing needs from times of low usage to peak demand. Some of the particular issues that are impacting the efficacy of onpremise grids include:

1

The expanding volume of data financial institutions need to gather, process, analyze, and store to run meaningful, accurate calculations.

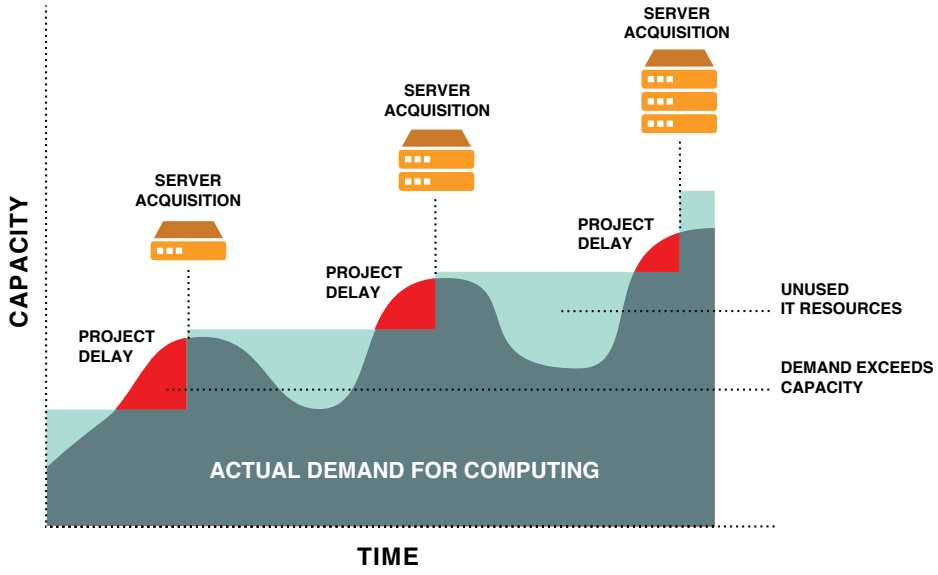
2

On-premise hardware flexibility is limited: since not all servers are optimized for grid-computing workloads, processing times often lag.

3

If markets experience volatility, regulations change, or a team wants to experiment with a new product, grid requirements often spike.

As graphic A illustrates, the only way an organization managing an on-premise grid can increase server capacity is by acquiring hardware, which is expensive and can delay production by weeks, if not months. Some organizations manage these capacity issues by adjusting calculations and workloads to work within the parameters of a limited grid. But such an approach forces tradeoffs between an environment that produces timely results with low overall utilization and an environment with high utilization that requires users to wait hours or days for results.



GRAPHIC A: The challenge of matching variable demand to on-premise capacity.

What if there were a better way?

Faced with the limitations of their on-premise storage and computing capabilities, Financial Services institutions of all sizes—FinTech startups, hedge funds, large insurers, and global investment banks, among others—are working with AWS to increase their grid-computing capabilities by extending their on-premise grids to the cloud or by building cloud-native grids. The elastic and on-demand nature of the AWS platform enables organizations to build and run grids when they are required. Unlike a traditional data center, where grid servers are powered up and available at all times, AWS instances can be shut down, saving significant procurement and maintenance costs.

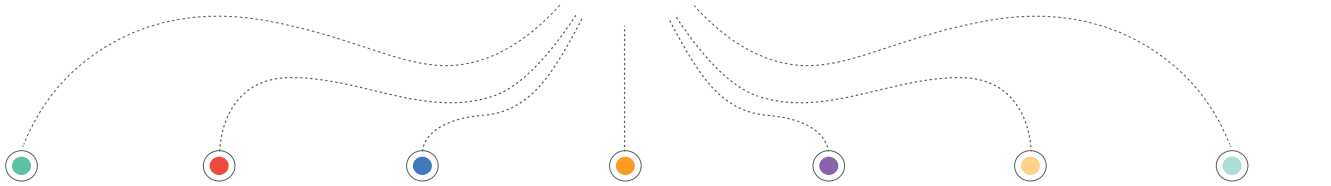
A common challenge Financial Services institutions are addressing by enhancing their existing grid infrastructure is a lack of adequate modelling capacity to assess the performance and risks related to trading operations. AWS recently worked with a leading hedge fund with a large quantitative research team to increase its grid capabilities by up to 75,000 cores at times of peak utilization. Before working with AWS, the compute-intensive nature of the team's models had led to bottlenecks and long job runs when researchers were testing their results, often at month end. Scaling up the grid helped the team more rapidly refine its trading strategies and reduce analysis time.

Benefits of a cloud approach

By building and running grids with AWS, companies are able to execute a larger number of parallel tasks, which leads to increased speed of analysis and reduced time to results. This flexibility is especially critical when moving into peak workload time periods. With AWS, as the need for compute resources arises, instances can automatically spin up to meet processing demands. After utilization has returned to the defined steady state and the "cooldown" period has been met, the system will automatically deprovision the additional resources. This elastic, cloudbased approach generates cost efficiencies, as users only run and pay for what they need. Organizations can also use it to scale their production, test, or development environments and alleviate the expense associated with disaster/backup-recovery environments.

AWS also offers a range of pricing models to reflect a variety of workloads. Customers can pay for capacity by the hour to address more ad-hoc needs, or they can lower costs further by reserving capacity for more extended durations. Organizations can even bid for unused capacity on a spot-price basis for less time-sensitive workloads. AWS further helps increase efficiency by providing as needed a range of computing types (CPU, GPU, and FPGA) optimized for specific applications, and without the need for large capital investments.

Why are Financial Services institutions increasingly shifting grid computing to the cloud?



VIRTUALLY UNLIMITED COMPUTE and storage resources

VARIOUS INSTANCE TYPES for specific types of workloads

COST OPTIMIZATION with a variety of pricing structures

ENHANCED SECURITY and compliance

EXPANDED BIG DATA CAPABILITIES for analysis and business intelligence

AUTOMATION CAPABILITIES for scaling and provisioning resources

MANAGEMENT CONSOLES to provide enterprise-wide control and visibility

Beyond scale: enhancing grids through the breadth of AWS' services

AWS provides a broad suite of services that can help organizations optimize their grid-computing workloads.

Virtually unlimited compute and storage

To further complement AWS' on-demand compute consumption model, AWS offers a complete range of cloud storage services to support both application and archival compliance requirements. Customers can opt for the scalable, durable object storage of **Amazon S3**, the high-performance file system of **Amazon FSx for Lustre**, the file system interface of **Amazon Elastic File System (EFS)**, or the persistent local storage of **Amazon Elastic Block Store (EBS)**. AWS also provides **Amazon Glacier**, a secure, and extremely low-cost cloud storage service for data archiving and long-term backup.

Various instance types for specific types of workloads

Amazon EC2 provides a wide selection of instance types optimized for different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and offer the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes one or more instance sizes, allowing you to scale your resources to the requirements of your target workload. Particularly relevant to grid workloads are: (1) C5 instances featuring the highest performing processors and the lowest price/compute performance; (2) X1 instances optimized for large-scale, in-memory applications; (3) P3 instances featuring GPU compute applications; and (4) z1d instances that provide both high compute performance and high memory, ideal for memory-intensive applications such as relational databases.

Cost optimization with a variety of pricing structures

There are multiple ways to access AWS instances, including **On Demand** (purchased by the hour with no commitments); **Reserved Instances** (reserved capacity assigned to a specific Availability Zone); **Scheduled Reserved Instances** (reserved capacity on a daily, weekly, or monthly recurring schedule over the course of a one-year term); **Spot Instances** (unused Amazon EC2 capacity purchased at auction); and **Spot Fleets** (easily managed fleet of Spot Instances that spans EC2 instance types and Availability Zones). Depending on a workload's flexibility in terms of start and end times, organizations can save up to 90% on On-Demand pricing by opting for Spot Instances. Reserved Instances also provide significant cost savings of up to 75% on On-Demand pricing.

Enhanced security and compliance

Security is job zero at AWS, and our customers benefit from a datacenter and network architecture built to meet the requirements of the most security-sensitive organizations. The **AWS Cloud Compliance Program** provides information on all the controls in place that demonstrate security assurance and meet the regulatory compliance requirements of the AWS platform and its respective users.

Expanded big data capabilities for analysis and business intelligence

Extracting insights and actionable information from data requires a broad array of technologies that can work with data efficiently, cost-effectively, and at scale. AWS offers the most comprehensive suite of big data services to handle every step of the analytics process chain, including data warehousing, business intelligence, batch processing, stream processing, machine learning, and data workflow orchestration.

Getting started with grid

Though many of the factors that are driving the need for increased grid computing capacity are common to most financial institutions, each organization's grid computing solution will ultimately depend on its specific challenges, its existing capabilities, and its unique technology environment. If you'd like to discuss your particular needs for increased computing power and flexibility, or if you have any questions regarding the benefits of a cloud-based approach, please don't hesitate to reach out to your AWS account manager. For additional materials, including a representative reference architecture and a demo video that provides insight into implementing a grid solution, please visit our website:

http://aws.amazon.com/financial-services/grid-computing